



## King's Research Portal

DOI:

[10.1093/biomet/asu075](https://doi.org/10.1093/biomet/asu075)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Doucet, A., Pitt, M., Deligiannidis, G., & Kohn, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *BIOMETRIKA*, 102(2), 295-313.  
<https://doi.org/10.1093/biomet/asu075>

### Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### Take down policy

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator

BY A. DOUCET

*Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, U.K.*  
doucet@stats.ox.ac.uk

M. K. PITT

*Department of Economics, University of Warwick, Coventry CV4 7AL, U.K.*  
m.pitt@warwick.ac.uk

G. DELIGIANNIDIS

*Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, U.K.*  
deligiannidis@stats.ox.ac.uk

AND R. KOHN

*Australian School of Business, University of New South Wales, Sydney 2052, Australia*  
r.kohn@unsw.edu.au

## SUMMARY

When an unbiased estimator of the likelihood is used within a Metropolis–Hastings chain, it is necessary to trade off the number of Monte Carlo samples used to construct this estimator against the asymptotic variances of the averages computed under this chain. Using many Monte Carlo samples will typically result in Metropolis–Hastings averages with lower asymptotic variances than the corresponding averages that use fewer samples; however, the computing time required to construct the likelihood estimator increases with the number of samples. Under the assumption that the distribution of the additive noise introduced by the loglikelihood estimator is Gaussian with variance inversely proportional to the number of samples and independent of the parameter value at which it is evaluated, we provide guidelines on the number of samples to select. We illustrate our results by considering a stochastic volatility model applied to stock index returns.

*Some key words:* Intractable likelihood; Metropolis–Hastings algorithm; Particle filter; Sequential Monte Carlo; State-space model.

## 1. INTRODUCTION

The use of unbiased estimators within the Metropolis–Hastings algorithm was initiated by [Lin et al. \(2000\)](#), with a surge of interest in these ideas since their introduction in Bayesian statistics by [Beaumont \(2003\)](#). In a Bayesian context, an unbiased likelihood estimator is commonly constructed using importance sampling as in [Beaumont \(2003\)](#) or particle filters as in [Andrieu et al. \(2010\)](#). [Andrieu & Roberts \(2009\)](#) call this method the pseudo-marginal algorithm and establish some of its theoretical properties.

Apart from the choice of proposals inherent to any Metropolis–Hastings algorithm, the main practical issue with the pseudo-marginal algorithm is the choice of the number,  $N$ , of Monte Carlo samples or particles used to estimate the likelihood. For any fixed  $N$ , the transition kernel of the pseudo-marginal algorithm leaves the posterior distribution of interest invariant. Using many Monte Carlo samples usually results in pseudo-marginal averages with lower asymptotic variances than the corresponding averages obtained from fewer samples, as established in a 2014 unpublished paper by C. Andrieu and M. Vihola (arXiv:1404.6909) for likelihood estimators based on importance sampling. Empirical evidence suggests that this result also holds when the likelihood is estimated by particle filters. However, the computing cost of constructing the likelihood estimator increases with  $N$ . We aim to select  $N$  so as to minimize the computational resources necessary to achieve a specified asymptotic variance for a particular pseudo-marginal average. This amount of computational resources, a quantity referred to as the computing time, is typically proportional to  $N$  times the asymptotic variance of the particular average, this variance being itself a function of  $N$ . Assuming that the distribution of the additive noise introduced by the loglikelihood estimator is Gaussian with a variance inversely proportional to  $N$  and independent of the parameter value at which it is evaluated, this minimization was carried out by Pitt et al. (2012) and Sherlock et al. (2015). However, Pitt et al. (2012) assumed that the Metropolis–Hastings proposal is the posterior density, whereas Sherlock et al. (2015) relaxed the Gaussian noise assumption but restricted themselves to an isotropic normal random walk proposal and assumed that the posterior density factorizes into  $d$  independent and identically distributed components, where  $d \rightarrow \infty$ .

Our article addresses a similar problem but considers general proposal and target densities and relaxes the Gaussian noise assumption. In this more general setting, we cannot minimize the computing time itself, and instead minimize explicit upper bounds on it. Quantitative results are presented under a Gaussian assumption. In this scenario, our guidelines are that  $N$  should be chosen such that the standard deviation of the loglikelihood estimator is around 1.0 when the Metropolis–Hastings algorithm using the exact likelihood is efficient and around 1.7 when it is inefficient. In most practical scenarios, the efficiency of the Metropolis–Hastings algorithm using the exact likelihood is unknown, as the algorithm cannot be implemented. In such cases, our results suggest selecting a standard deviation around 1.2.

## 2. METROPOLIS–HASTINGS METHOD USING AN ESTIMATED LIKELIHOOD

We briefly review how an unbiased likelihood estimator can be used within a Metropolis–Hastings scheme in a Bayesian context. Let  $y \in \mathcal{Y}$  be the observations and  $\theta \in \Theta \subseteq \mathbb{R}^d$  the parameters of interest. The likelihood is denoted by  $p(y | \theta)$ , and the prior for  $\theta$  admits a density  $p(\theta)$  with respect to Lebesgue measure, so the posterior density of interest is  $\pi(\theta) \propto p(y | \theta)p(\theta)$ . We slightly abuse notation by using the same symbols for distributions and densities.

The Metropolis–Hastings scheme to sample from  $\pi$  simulates a Markov chain according to the transition kernel

$$Q_{\text{EX}}(\theta, d\vartheta) = q(\theta, \vartheta)\alpha_{\text{EX}}(\theta, \vartheta) d\vartheta + \{1 - \varrho_{\text{EX}}(\theta)\}\delta_{\theta}(d\vartheta),$$

where

$$\alpha_{\text{EX}}(\theta, \vartheta) = \min\{1, r_{\text{EX}}(\theta, \vartheta)\}, \quad \varrho_{\text{EX}}(\theta) = \int q(\theta, \vartheta)\alpha_{\text{EX}}(\theta, \vartheta) d\vartheta,$$

with  $r_{\text{EX}}(\theta, \vartheta) = \pi(\vartheta)q(\vartheta, \theta)/\{\pi(\theta)q(\theta, \vartheta)\}$ . This Markov chain cannot be simulated if  $p(y | \theta)$  is intractable.

Assume that  $p(y | \theta)$  is intractable but we have access to a nonnegative unbiased estimator  $\hat{p}(y | \theta, U)$  of  $p(y | \theta)$ , where  $U \sim m(\cdot)$  represents all the auxiliary random variables used to obtain this estimator. In this case, we introduce the joint density  $\bar{\pi}(\theta, u)$  on  $\Theta \times \mathcal{U}$ , where

$$\bar{\pi}(\theta, u) = \pi(\theta)m(u)\hat{p}(y | \theta, u)/p(y | \theta). \quad (1)$$

This joint density admits the correct marginal density  $\pi(\theta)$ , because  $\hat{p}(y | \theta, U)$  is unbiased. The pseudo-marginal algorithm is a Metropolis–Hastings scheme targeting (1) with proposal density  $q(\theta, \cdot)m(\cdot)$ , yielding the acceptance probability

$$\min \left\{ 1, \frac{\hat{p}(y | \vartheta, v)p(\vartheta)q(\vartheta, \theta)}{\hat{p}(y | \theta, u)p(\theta)q(\theta, \vartheta)} \right\} = \min \left\{ 1, \frac{\hat{p}(y | \vartheta, v)/p(y | \vartheta)}{\hat{p}(y | \theta, u)/p(y | \theta)} r_{\text{EX}}(\theta, \vartheta) \right\} \quad (2)$$

for a proposal  $(\vartheta, v)$ . In practice, we record only  $\{\theta, \log \hat{p}(y | \theta, u)\}$  instead of  $(\theta, u)$ . We follow [Andrieu & Roberts \(2009\)](#) and [Pitt et al. \(2012\)](#) and analyse this scheme using additive noise,  $Z = \log \hat{p}(y | \theta, U) - \log p(y | \theta) = \psi(\theta, U)$ , in the loglikelihood estimator, rather than  $U$ . In this parameterization, the target density on  $\Theta \times \mathbb{R}$  becomes

$$\bar{\pi}(\theta, z) = \pi(\theta) \exp(z)g(z | \theta), \quad (3)$$

where  $g(z | \theta)$  is the density of  $Z$  when  $U \sim m(\cdot)$  and the transformation  $Z = \psi(\theta, U)$  is applied.

To sample from  $\bar{\pi}(\theta, z)$ , we could use the scheme previously described to sample from  $\bar{\pi}(\theta, u)$  and then set  $z = \psi(\theta, u)$ . Equivalently, we can use the transition kernel

$$\begin{aligned} Q\{(\theta, z), (d\vartheta, dw)\} &= q(\theta, \vartheta)g(w | \vartheta)\alpha_Q\{(\theta, z), (\vartheta, w)\} d\vartheta dw \\ &\quad + \{1 - \varrho_Q(\theta, z)\}\delta_{(\theta, z)}(d\vartheta, dw), \end{aligned} \quad (4)$$

where

$$\alpha_Q\{(\theta, z), (\vartheta, w)\} = \min\{1, \exp(w - z) r_{\text{EX}}(\theta, \vartheta)\} \quad (5)$$

is (2) expressed in the new parameterization. Henceforth, we make the following assumption.

*Assumption 1.* The noise density  $g(z)$  is independent of  $\theta$ .

Under this assumption, the target density (3) factorizes as  $\pi(\theta)\pi_z(z)$ , where

$$\pi_z(z) = \exp(z)g(z). \quad (6)$$

Assumption 1 allows us to analyse the performance of the pseudo-marginal algorithm in detail, but it is not satisfied in practical settings. However, in the stationary regime, we are concerned with the noise density at values of the parameter which arise from the target density  $\pi(\theta)$  and the marginal density of the proposals at stationarity,  $\int \pi(d\vartheta)q(\vartheta, \theta)$ . If the noise density does not vary significantly in regions of high probability mass of these densities, then this assumption is approximately satisfied. In §4, we examine experimentally how the noise density varies with respect to draws from  $\pi(\theta)$  and  $\int \pi(d\vartheta)q(\vartheta, \theta)$ .

## 3. MAIN RESULTS

## 3.1. Outline

This section presents the main results of the paper. The proofs are given in Appendix 1 and the Supplementary Material. We minimize upper bounds on the computing time of the pseudo-marginal algorithm, as discussed in § 1. This requires us to establish upper bounds on the asymptotic variance of an ergodic average under the kernel  $Q$  given in (4). To obtain these bounds, we introduce a new Markov kernel  $Q^*$ , defined by

$$Q^*\{(\theta, z), (d\vartheta, dw)\} = q(\theta, \vartheta)g(w)\alpha_{Q^*}\{(\theta, z), (\vartheta, w)\} d\vartheta dw \\ + \{1 - \varrho_{\text{EX}}(\theta)\varrho_Z(z)\}\delta_{(\theta, z)}(d\vartheta, dw),$$

where

$$\alpha_{Q^*}\{(\theta, z), (\vartheta, w)\} = \alpha_{\text{EX}}(\theta, \vartheta)\alpha_Z(z, w), \quad \alpha_Z(z, w) = \min\{1, \exp(w - z)\}, \quad (7)$$

$$\varrho_Z(z) = \int g(w)\alpha_Z(z, w) dw.$$

As  $Q$  and  $Q^*$  are reversible with respect to  $\bar{\pi}$  and the acceptance probability in (7) is always smaller than (5), an application of the theorem in [Peskun \(1973\)](#) ensures that the variance of an ergodic average under  $Q^*$  is greater than or equal to the variance under  $Q$ . We obtain an exact expression for the variance under the bounding kernel  $Q^*$ , as well as simpler upper bounds, by exploiting a nonstandard representation of this variance, the factor form of the acceptance probability in (7), and the spectral properties of an auxiliary Markov kernel.

## 3.2. Inefficiency of Metropolis–Hastings-type chains

In this subsection we recall and establish various results on the integrated autocorrelation time of Markov chains, henceforth referred to as the inefficiency. In particular, we give a novel representation of the inefficiency of Metropolis–Hastings-type chains, which will be the basic ingredient of the proof of our main result.

Consider a Markov kernel  $\Pi$  on the measurable space  $(\mathbf{X}, \mathcal{X}) = \{\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\}$ , where  $\mathcal{B}(\mathbb{R}^n)$  is the Borel  $\sigma$ -algebra on  $\mathbb{R}^n$ . For any measurable real-valued function  $f$ , measurable set  $A$  and probability measure  $\mu$ , we write  $\mu(f) = \int_{\mathbf{X}} \mu(dx) f(x)$ ,  $\bar{f} = f - \mu(f)$ ,  $\mu(A) = \mu\{\mathbb{I}_A(\cdot)\}$ ,  $\Pi f(x) = \int_{\mathbf{X}} \Pi(x, dy) f(y)$  and, for  $n \geq 2$ ,  $\Pi^n(x, dy) = \int_{\mathbf{X}} \Pi^{n-1}(x, dz) \Pi(z, dy)$ , with  $\Pi^1 = \Pi$ . We introduce the Hilbert spaces

$$L^2(\mathbf{X}, \mu) = \{f: \mathbf{X} \rightarrow \mathbb{R} : \mu(f^2) < \infty\}, \quad L_0^2(\mathbf{X}, \mu) = \{f: \mathbf{X} \rightarrow \mathbb{R} : \mu(f) = 0, \mu(f^2) < \infty\}$$

equipped with the inner product  $\langle f, g \rangle_{\mu} = \int f(x)g(x)\mu(dx)$ . We write  $\phi_n(f, \Pi) = \langle \bar{f}, \Pi^n \bar{f} \rangle_{\mu} / \mu(\bar{f}^2)$  for the autocorrelation at lag  $n \geq 0$  and  $\text{IF}(f, \Pi) = 1 + 2 \sum_{n=1}^{\infty} \phi_n(f, \Pi)$  for the associated inefficiency. A  $\mu$ -invariant and  $\psi$ -irreducible Markov chain is said to be ergodic; see [Tierney \(1994\)](#) for the definition of  $\psi$ -irreducibility. The next result follows directly from [Kipnis & Varadhan \(1986\)](#) and [Häggström & Rosenthal \(2007, Theorem 4 and Corollary 6\)](#).

**PROPOSITION 1.** *Consider a  $\mu$ -reversible and ergodic Markov kernel  $\Pi$ . Let  $(X_i)_{i \geq 1}$  be a stationary Markov chain that evolves according to  $\Pi$ , and let  $h \in L^2(\mathbf{X}, \mu)$  be such that  $\mu(\bar{h}^2) > 0$ . Then the following properties hold.*

- (i) There exists a probability measure  $e(h, \Pi)$  on  $[-1, 1)$ , called the spectral measure, such that

$$\phi_n(h, \Pi) = \int_{-1}^1 \lambda^n e(h, \Pi)(d\lambda), \quad \text{IF}(h, \Pi) = \int_{-1}^1 (1 + \lambda)(1 - \lambda)^{-1} e(h, \Pi)(d\lambda). \quad (8)$$

- (ii) If  $\text{IF}(h, \Pi) < \infty$ , then as  $n \rightarrow \infty$ ,

$$n^{-1/2} \sum_{i=1}^n \{h(X_i) - \mu(h)\} \rightarrow N\{0, \mu(\tilde{h}^2)_{\text{IF}(h, \Pi)}\} \quad (9)$$

in distribution, where  $N(a, b^2)$  denotes the normal distribution with mean  $a$  and variance  $b^2$ .

When estimating  $\mu(h)$ , (9) implies that we need approximately  $n_{\text{IF}(h, \Pi)}$  samples from the Markov chain  $(X_i)_{i \geq 1}$  to obtain an estimator of the same precision as an average of  $n$  independent draws from  $\mu$ .

Henceforth we consider a  $\mu$ -reversible kernel

$$P(x, dy) = q(x, dy)\alpha(x, y) + \{1 - \varrho(x)\}\delta_x(dy), \quad \varrho(x) = \int q(x, dy)\alpha(x, y),$$

where the proposal kernel is selected such that  $q(x, \{x\}) = 0$ ,  $\alpha(x, y)$  is the acceptance probability, and we assume that there does not exist an  $x$  such that  $\mu(\{x\}) = 1$ . We refer to  $P$  as a Metropolis–Hastings-type kernel because it is structurally similar to the Metropolis–Hastings kernel but we do not require  $\alpha(x, y)$  to be the Metropolis–Hastings acceptance probability. This generalization is required when studying the kernel  $Q^*$ , as the acceptance probability  $\alpha_{Q^*}\{(\theta, z), (\vartheta, w)\}$  in (7) is not the Metropolis–Hastings acceptance probability.

Let  $(X_i)_{i \geq 1}$  be a Markov chain that evolves according to  $P$ . We now establish a nonstandard expression for  $\text{IF}(h, P)$  derived from the associated jump chain representation  $(\tilde{X}_i, \tau_i)_{i \geq 1}$  of  $(X_i)_{i \geq 1}$ . In this representation,  $(\tilde{X}_i)_{i \geq 1}$  corresponds to the sequence of accepted proposals and  $(\tau_i)_{i \geq 1}$  the associated sojourn times, i.e.,  $\tilde{X}_1 = X_1 = \dots = X_{\tau_1}$ ,  $\tilde{X}_2 = X_{\tau_1+1} = \dots = X_{\tau_1+\tau_2}$  and so on, with  $\tilde{X}_{i+1} \neq \tilde{X}_i$ . We now state some properties of this jump chain; see Lemma 1 in Douc & Robert (2011).

LEMMA 1. Let  $P$  be  $\psi$ -irreducible. Then  $\varrho(x) > 0$  for any  $x \in \mathbf{X}$  and  $(\tilde{X}_i, \tau_i)_{i \geq 1}$  is a Markov chain with a  $\bar{\mu}$ -reversible transition kernel  $\bar{P}$ , where

$$\bar{P}\{(x, \tau), (dy, \zeta)\} = \tilde{P}(x, dy)G\{\zeta; \varrho(y)\}, \quad \bar{\mu}(dx, \tau) = \tilde{\mu}(dx)G\{\tau; \varrho(x)\}$$

with

$$\tilde{P}(x, dy) = \frac{q(x, dy)\alpha(x, y)}{\varrho(x)}, \quad \tilde{\mu}(dx) = \frac{\mu(dx)\varrho(x)}{\mu(\varrho)},$$

and  $G(\cdot; \nu)$  denotes the geometric distribution with parameter  $\nu$ .

The next proposition gives the relationship between  $\text{IF}(h, P)$  and  $\text{IF}(h/\varrho, \tilde{P})$ .

PROPOSITION 2. Assume that  $P$  and  $\tilde{P}$  are ergodic, that  $h \in L_0^2(\mathbf{X}, \mu)$  and that  $\text{IF}(h, P) < \infty$ . Then  $h/\varrho \in L_0^2(\mathbf{X}, \tilde{\mu})$ ,

$$\mu(h^2)\{1 + \text{IF}(h, P)\} = \mu(\varrho)\tilde{\mu}(h^2/\varrho^2)\{1 + \text{IF}(h/\varrho, \tilde{P})\}$$

and  $\text{IF}(h/\varrho, \tilde{P}) \leq \text{IF}(h, P)$ .

Lemma 1 and Proposition 2 are used in § 3.3 to establish a representation of the inefficiency for the kernel  $P = Q^*$ .

We conclude this section by establishing some results on the positivity of the Metropolis–Hastings kernel and its associated jump kernel. Recall that a  $\mu$ -invariant Markov kernel  $\Pi$  is positive if  $\langle \Pi h, h \rangle_\mu \geq 0$  for any  $h \in L^2(\mathbf{X}, \mu)$ . If  $\Pi$  is reversible, then positivity is equivalent to  $e(h, \Pi)([0, 1]) = 1$  for all  $h \in L^2(\mathbf{X}, \mu)$ , where  $e(h, \Pi)$  is the spectral measure, and positivity also implies that  $\text{IF}(h, \Pi) \geq 1$ ; see, for example, Geyer (1992). The positivity of the jump kernel  $\tilde{P}$  associated with a Metropolis–Hastings kernel  $P$  is useful here, as several bounds on the inefficiency established subsequently require the spectral measure  $e(h, \tilde{P})$  to be supported on  $[0, 1]$ . We now give sufficient conditions ensuring this property by extending Lemma 3.1 of Baxendale (2005). This complements results of Rudolf & Ullrich (2013).

PROPOSITION 3. Assume that  $\alpha(x, y)$  is the Metropolis–Hastings acceptance probability and  $\mu(\mathrm{d}x) = \mu(x) \mathrm{d}x$ . If  $P$  is  $\psi$ -irreducible, then  $\tilde{P}$  and  $P$  are both positive if one of the following two conditions is satisfied:

- (i)  $q(x, \mathrm{d}y) = q(x, y) \mathrm{d}y$  is a  $v$ -reversible kernel with  $v(\mathrm{d}x) = v(x) \mathrm{d}x$ ,  $\mu$  is absolutely continuous with respect to  $v$ , and there exists  $r : \mathbf{X} \times \mathbf{Z} \rightarrow \mathbb{R}^+$  such that  $v(x)q(x, y) = \int r(x, z)r(y, z)\chi(\mathrm{d}z)$ , where  $\chi$  is a measure on  $\mathbf{Z}$ ;
- (ii)  $q(x, \mathrm{d}y) = q(x, y) \mathrm{d}y$ , and there exists  $s : \mathbf{X} \times \mathbf{Z} \rightarrow \mathbb{R}^+$  such that  $q(x, y) = \int s(x, z)s(y, z)\chi(\mathrm{d}z)$ , where  $\chi$  is a measure on  $\mathbf{Z}$ .

Remark 1. Condition (i) in the proposition is satisfied for an independent proposal  $q(x, y) = v(y)$  by taking  $\mathbf{Z} = \{1\}$ ,  $\chi(\mathrm{d}z) = \delta_1(\mathrm{d}z)$  and  $r(x, 1) = v(x)$ . It is also satisfied for autoregressive positively correlated proposals with normal or Student- $t$  innovations. Condition (ii) holds if  $q(x, y)$  is a symmetric random walk proposal whose increments are multivariate normal or Student- $t$ .

### 3.3. Inefficiency of the bounding chain

In this subsection we apply the results of § 3.2 to establish an exact expression for  $\text{IF}(h, Q^*)$ . The next lemma shows that  $\text{IF}(h, Q^*)$  is an upper bound on  $\text{IF}(h, Q)$ .

LEMMA 2. The kernel  $Q^*$  is  $\bar{\pi}$ -reversible and  $\text{IF}(h, Q) \leq \text{IF}(h, Q^*)$  for any  $h \in L^2(\Theta \times \mathbb{R}, \bar{\pi})$ .

In practice, we are only interested in functions  $h \in L^2(\Theta, \pi)$ . To simplify notation, we write  $\text{IF}(h, Q)$  in this case, instead of introducing the function  $\tilde{h} \in L^2(\Theta \times \mathbb{R}, \bar{\pi})$  which satisfies  $\tilde{h}(\theta, z) = h(\theta)$  for all  $z \in \mathbb{R}$  and writing  $\text{IF}(\tilde{h}, Q)$ . Proposition 2 shows that it is possible to express  $\text{IF}(h, Q^*)$  as a function of the inefficiency of its jump kernel  $\tilde{Q}^*$ , which is particularly useful as  $\tilde{Q}^*$  admits a simple structure.



LEMMA 3. Assume that  $Q^*$  is  $\tilde{\pi}$ -irreducible. The jump kernel  $\tilde{Q}^*$  associated with  $Q^*$  is

$$\tilde{Q}^*\{(\theta, z), (d\vartheta, dw)\} = \tilde{Q}_{\text{EX}}(\theta, d\vartheta) \tilde{Q}_z(z, dw), \quad (10)$$

where

$$\tilde{Q}_{\text{EX}}(\theta, d\vartheta) = \frac{q(\theta, \vartheta) \alpha_{\text{EX}}(\theta, \vartheta) d\vartheta}{Q_{\text{EX}}(\theta)}, \quad \tilde{Q}_z(z, dw) = \frac{g(w) \alpha_z(z, w) dw}{Q_z(z)}.$$

The kernel  $\tilde{Q}_{\text{EX}}$  is reversible with respect to  $\tilde{\pi}(d\theta)$ , and the kernel  $\tilde{Q}_z$  is positive and reversible with respect to  $\tilde{\pi}_z(dz)$ , where

$$\tilde{\pi}(d\theta) = \frac{\pi(d\theta) Q_{\text{EX}}(\theta)}{\pi(Q_{\text{EX}})}, \quad \tilde{\pi}_z(dz) = \frac{\pi_z(dz) Q_z(z)}{\pi_z(Q_z)}.$$

Moreover,  $\tilde{Q}_z$  is uniformly ergodic and if  $\pi_z(1/Q_z) < \infty$ , then  $\text{IF}(1/Q_z, \tilde{Q}_z) < \infty$ .

In addition, if  $Q^*$  is ergodic,  $h \in L_0^2(\Theta, \pi)$ ,  $\text{IF}(h, Q^*) < \infty$  and  $\tilde{Q}^*$  is ergodic, then  $h/Q_{\text{EX}} \in L_0^2(\Theta, \tilde{\pi})$ ,  $\pi_z(1/Q_z) < \infty$ ,  $\text{IF}\{h/(Q_{\text{EX}}Q_z), \tilde{Q}^*\} < \infty$  and

$$\pi(h^2)\{1 + \text{IF}(h, Q^*)\} = \pi(Q_{\text{EX}})\pi_z(1/Q_z)\tilde{\pi}(h^2/Q_{\text{EX}}^2)[1 + \text{IF}\{h/(Q_{\text{EX}}Q_z), \tilde{Q}^*\}]. \quad (11)$$

The following theorem provides an expression for  $\text{IF}(h, Q^*)$  which decouples the contributions of the parameter and the noise components. The proof exploits the relationships between  $\text{IF}(h, Q_{\text{EX}})$  and  $\text{IF}(h/Q_{\text{EX}}, \tilde{Q}_{\text{EX}})$  and between  $\text{IF}(h, Q^*)$  and  $\text{IF}\{h/(Q_{\text{EX}}Q_z), \tilde{Q}^*\}$ , as well as the spectral representation (8) of  $\text{IF}\{h/(Q_{\text{EX}}Q_z), \tilde{Q}^*\}$ . This spectral representation admits a simple structure due to the product form (10) of  $\tilde{Q}^*$ .

THEOREM 1. Let  $h \in L^2(\Theta, \pi)$ . Assume that  $Q_{\text{EX}}$ ,  $Q^*$ ,  $\tilde{Q}_{\text{EX}}$  and  $\tilde{Q}^*$  are ergodic with  $\text{IF}(h, Q^*) < \infty$ . Then  $\text{IF}(h, Q) \leq \text{IF}(h, Q^*)$  and  $\text{IF}(h, Q^*)$  equals

$$\begin{aligned} & \frac{1 + \text{IF}(h, Q_{\text{EX}})}{\pi_z(Q_z)} - 1 \\ & + \frac{2\{1 + \text{IF}(h, Q_{\text{EX}})\}}{1 + \text{IF}(h/Q_{\text{EX}}, \tilde{Q}_{\text{EX}})} \left\{ \pi_z(1/Q_z) - \frac{1}{\pi_z(Q_z)} \right\} \sum_{n=0}^{\infty} \phi_n(h/Q_{\text{EX}}, \tilde{Q}_{\text{EX}}) \phi_n(1/Q_z, \tilde{Q}_z). \end{aligned}$$

Remark 2. If  $q(\theta, \vartheta) = \pi(\vartheta)$ , then  $\text{IF}(h, Q_{\text{EX}}) = \text{IF}(h/Q_{\text{EX}}, \tilde{Q}_{\text{EX}}) = 1$  and  $\phi_n(h/Q_{\text{EX}}, \tilde{Q}_{\text{EX}}) = 0$  for  $n \geq 1$ . It follows from Theorem 1 that  $\text{IF}(h, Q^*) = 2\pi_z(1/Q_z) - 1$ . This result was established in Lemma 4 of Pitt et al. (2012).

Theorem 1 requires  $Q_{\text{EX}}$ ,  $Q^*$ ,  $\tilde{Q}_{\text{EX}}$  and  $\tilde{Q}^*$  to be ergodic. The following proposition, which generalizes Theorem 2.2 of Roberts & Tweedie (1996), provides sufficient conditions ensuring this.

PROPOSITION 4. Suppose that  $\pi(\theta)$  is bounded away from 0 and  $\infty$  on compact sets, and that there exist  $\delta > 0$  and  $\varepsilon > 0$  such that  $|\theta - \vartheta| \leq \delta$  implies  $q(\theta, \vartheta) \geq \varepsilon$ . Then  $Q_{\text{EX}}$ ,  $Q^*$ ,  $\tilde{Q}_{\text{EX}}$  and  $\tilde{Q}^*$  are ergodic.

### 3.4. Bounds on the relative inefficiency of the pseudo-marginal chain

For any kernel  $\Pi$ , we define the relative inefficiency  $\text{RIF}(h, \Pi) = \text{IF}(h, \Pi)/\text{IF}(h, Q_{\text{EX}})$ , which measures the inefficiency of  $\Pi$  compared to that of  $Q_{\text{EX}}$ . This section provides tractable upper bounds for  $\text{RIF}(h, Q)$ . From Lemma 2,  $\text{RIF}(h, Q) \leq \text{RIF}(h, Q^*)$ , but the expression of



$\text{RIF}(h, Q^*)$  that follows from Theorem 1 is intricate and depends on the autocorrelation sequence  $\{\phi_n(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}_{n \geq 1}$  as well as on other terms. The next corollary provides upper bounds on  $\text{RIF}(h, Q)$  that depend only on  $\text{IF}(h, Q_{\text{EX}})$  and  $\phi_1(1/\varrho_Z, \tilde{Q}_Z)$ . The latter can be evaluated numerically and, to simplify the notation, we refer to it as  $\phi_Z$ .

COROLLARY 1. *Under the assumptions of Theorem 1, the following hold:*

(i)  $\text{RIF}(h, Q) \leq \text{URIF}_1(h)$ , where

$$\text{URIF}_1(h) = \{1 + 1/\text{IF}(h, Q_{\text{EX}})\}[\pi_Z(1/\varrho_Z) + (1 - \phi_Z)\{\pi_Z(1/\varrho_Z) - 1/\pi_Z(Q_Z)\}] - 1/\text{IF}(h, Q_{\text{EX}});$$

(ii) if, in addition,  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \geq 1$ , then  $\text{RIF}(h, Q) \leq \text{URIF}_2(h) \leq \text{URIF}_1(h)$ , where

$$\text{URIF}_2(h) = \{1 + 1/\text{IF}(h, Q_{\text{EX}})\}\pi_Z(1/\varrho_Z) - 1/\text{IF}(h, Q_{\text{EX}}).$$

Proposition 3 gives sufficient conditions for the condition  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \geq 1$  of Corollary 1(ii) to hold.

*Remark 3.* The bounds above are tight in two cases. First, if  $\pi_Z(1/\varrho_Z) \rightarrow 1$ , then  $\text{RIF}(h, Q), \text{URIF}_1(h), \text{URIF}_2(h) \rightarrow 1$ . Second, if  $q(\theta, \vartheta) = \pi(\vartheta)$ , then  $\text{RIF}(h, Q) = \text{URIF}_2(h)$ .

We now provide upper bounds on  $\text{RIF}(h, Q)$  and lower bounds on  $\text{RIF}(h, Q^*)$  in terms of  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$ .

COROLLARY 2. *Under the assumptions of Theorem 1:*

(i)  $\text{RIF}(h, Q) \leq \text{URIF}_3(h)$  where

$$\begin{aligned} \text{URIF}_3(h) = & \left\{1 + \frac{1}{\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})}\right\} \left[ \frac{1}{\pi_Z(Q_Z)} + \phi_Z \left\{ \pi_Z(1/\varrho_Z) - \frac{1}{\pi_Z(Q_Z)} \right\} \right] \\ & + 2\{\pi_Z(1/\varrho_Z) - 1/\pi_Z(Q_Z)\}(1 - \phi_Z)/\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 1/\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}); \end{aligned}$$

(ii)  $\text{RIF}(h, Q) \leq \text{URIF}_4(h)$  where

$$\begin{aligned} \text{URIF}_4(h) = & \frac{\{1 + 1/\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \{\pi_Z(1/\varrho_Z) - 1/\pi_Z(Q_Z)\} \{1 + \text{IF}(1/\varrho_Z, \tilde{Q}_Z)\} \\ & + 1/\pi_Z(Q_Z) + \frac{1}{\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \left\{ \frac{1}{\pi_Z(Q_Z)} - 1 \right\}; \end{aligned}$$

(iii) if  $\tilde{Q}_{\text{EX}}$  is positive, then  $\text{RIF}(h, Q^*) \geq \text{LRIF}_1(h)$  where

$$\text{LRIF}_1(h) = \frac{1}{\pi_Z(Q_Z)} + \frac{2}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \{\pi_Z(1/\varrho_Z) - 1/\pi_Z(Q_Z)\};$$

(iv)  $\text{RIF}(h, Q^*) \geq \text{LRIF}_2$  where

$$\text{LRIF}_2 = 1/\pi_Z(Q_Z), \tag{12}$$

and  $\text{RIF}(h, Q^*), \text{URIF}_4(h) \rightarrow \text{LRIF}_2$  as  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \rightarrow \infty$ .

Again, Proposition 3 gives sufficient conditions for  $\tilde{Q}_{\text{EX}}$  to be positive. In § 3.5 these bounds are discussed in more detail.

### 3.5. Optimizing the computing time under a Gaussian assumption

This subsection provides quantitative guidelines on how to select the standard deviation  $\sigma$  of the noise density, under the following assumption.

*Assumption 2.* The noise density is  $g^\sigma(z) = \varphi(z; -\sigma^2/2, \sigma^2)$ , where  $\varphi(z; a, b^2)$  is a univariate normal density with mean  $a$  and variance  $b^2$ .

Assumption 2 ensures that  $\int \exp(z)g^\sigma(z) dz = 1$ , as required by the unbiasedness of the likelihood estimator. Consider a time series  $y_{1:T} = (y_1, \dots, y_T)$ , where the likelihood estimator  $\hat{p}(y_{1:T} | \theta)$  of  $p(y_{1:T} | \theta)$  is computed through a particle filter with  $N$  particles. Theorem 1 of [Bérard et al. \(2014\)](#) shows that, under regularity assumptions, the loglikelihood error is distributed according to a normal density with mean  $-\delta\gamma^2/2$  and variance  $\delta\gamma^2$  as  $T \rightarrow \infty$ , for  $N = \delta^{-1}T$ . Hence, in this important scenario, the noise distribution satisfies approximately the form specified in Assumption 2 for large  $T$ , and the variance is asymptotically inversely proportional to the number of samples. This assumption is also made in [Pitt et al. \(2012\)](#), where it is justified experimentally. Section 4 below provides additional experimental results.

The next result is Lemma 4 in [Pitt et al. \(2012\)](#) and follows from Assumption 2, equation (6) and Remark 2. We now make the dependence on  $\sigma$  explicit in our notation.

**COROLLARY 3.** Under Assumption 2,  $\pi_z^\sigma(z) = \varphi(z; \sigma^2/2, \sigma^2)$ ,

$$\begin{aligned} \varrho_z^\sigma(z) &= 1 - \Phi(z/\sigma + \sigma/2) + \exp(-z)\Phi(z/\sigma - \sigma/2), \\ \pi_z^\sigma(1/\varrho_z^\sigma) &= \int \frac{\varphi(w; 0, 1)}{1 - \bar{\varrho}_z^\sigma(w)} dw, \end{aligned}$$

where  $\bar{\varrho}_z^\sigma(w) = \Phi(w + \sigma) - \exp(-w\sigma - \sigma^2/2)\Phi(w)$  and  $\Phi(\cdot)$  is the standard Gaussian cumulative distribution function. Additionally,  $\pi_z^\sigma(\varrho_z^\sigma) = 2\Phi(-\sigma/\sqrt{2})$ .

The terms  $\pi_z^\sigma(1/\varrho_z^\sigma)$ ,  $\phi_z^\sigma$  and  $\text{IF}(1/\varrho_z^\sigma, \tilde{Q}_z)$  appearing in the bounds of Corollaries 1 and 2 do not admit analytical expressions, but they can be computed numerically. We note that  $\pi_z^\sigma(1/\varrho_z^\sigma)$  is finite and thus, by Lemma 3,  $\text{IF}(1/\varrho_z^\sigma, \tilde{Q}_z)$  is also finite. Consequently, for specific values of  $\sigma$ ,  $\text{IF}(h, Q_{\text{EX}})$  and  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$ , these bounds can be calculated.

We now use these bounds to guide the choice of  $\sigma$ . The quantity we aim to minimize is the relative computing time for  $Q$ , defined as  $\text{RCT}(h, Q; \sigma) = \text{RIF}(h, Q; \sigma)/\sigma^2$ , because  $1/\sigma^2$  is usually approximately proportional to the number of samples  $N$  used to estimate the likelihood and the computational cost at each iteration is typically proportional to  $N$ , at least in the particle filter scenario described previously. We define  $\text{RCT}(h, Q^*; \sigma)$  similarly. As  $\text{RIF}(h, Q; \sigma)$  is intractable, we instead minimize the upper bounds  $\text{URCT}_i(h; \sigma) = \text{URIF}_i(h; \sigma)/\sigma^2$  for  $i = 1, \dots, 4$ . We similarly define the quantities  $\text{LRCT}_1(h; \sigma) = \text{LRIF}_1(h; \sigma)/\sigma^2$  and  $\text{LRCT}_2(\sigma) = \text{LRIF}_2(\sigma)/\sigma^2$ , which bound  $\text{RCT}(h, Q^*; \sigma)$  from below. Figure 1 plots these bounds, except  $\text{LRCT}_1(h; \sigma)$ , for clarity, against  $\sigma$  for different values of  $\text{IF}(h, Q_{\text{EX}})$  and  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$ . It is clear, however, that  $\text{LRCT}_1(h; \sigma)$  provides a much tighter lower bound than  $\text{LRCT}_2(\sigma)$  for low values of  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$ .

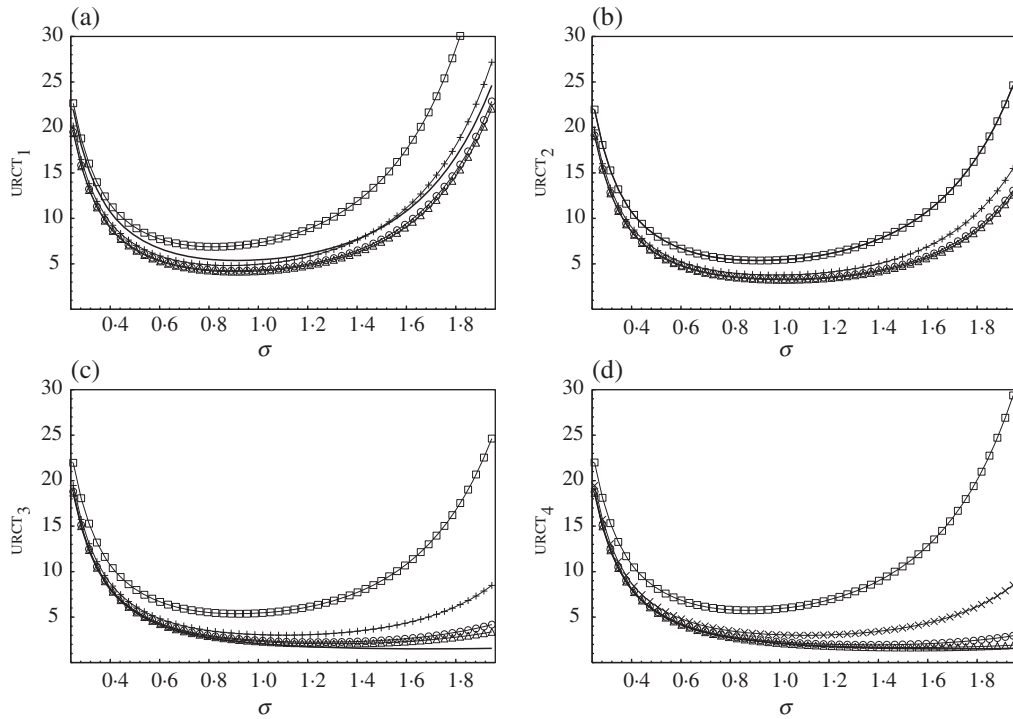


Fig. 1. Theoretical results for relative computing time as a function of  $\sigma$ : (a)  $\text{URCT}_1(h; \sigma)$  and (b)  $\text{URCT}_2(h; \sigma)$ , with values of  $\text{IF}(h, Q_{\text{EX}})$  taken to be 1 (squares), 4 (crosses), 20 (circles) and 80 (triangles), and where the solid line corresponds to the perfect proposal, as discussed in Remark 2; (c)  $\text{URCT}_3(h; \sigma)$  and (d)  $\text{URCT}_4(h; \sigma)$  plotted together with the lower bound  $\text{LRCT}_2(\sigma)$  (solid line), with values of  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$  taken to be 1 (squares), 4 (crosses), 20 (circles) and 80 (triangles).

Before discussing how these results guide the selection of  $\sigma$ , we outline some properties of the bounds. First, as the corresponding inefficiency increases, the upper bounds  $\text{URCT}_i(h; \sigma)$  displayed in Fig. 1 become flatter as functions of  $\sigma$ , and the corresponding minimizing argument  $\sigma_{\text{opt}}$  increases. This flattening effect suggests less sensitivity to the choice of  $\sigma$  for the pseudo-marginal algorithm. Second, for given  $\sigma$ , all the upper bounds are decreasing functions of the corresponding inefficiency, which suggests that the penalty from using the pseudo-marginal algorithm drops as the exact algorithm becomes more inefficient. Third, in the case discussed in Remark 2, where  $q(\theta, \vartheta) = \pi(\vartheta)$  so that  $\text{IF}(h, Q_{\text{EX}}) = \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) = 1$ , we obtain  $\text{URCT}_2(h; \sigma) = \text{URCT}_3(h; \sigma) = \text{RCT}(h, Q^*; \sigma) = \text{RCT}(h, Q; \sigma)$ . Fourth,  $\text{URCT}_4(h; \sigma)$  agrees with the lower bound  $\text{LRCT}_2(\sigma)$  as  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \rightarrow \infty$ , as indicated by Corollary 2(ii). In this case these two bounds, as well as  $\text{URCT}_1(h; \sigma)$ , are sharp for  $\text{RCT}(h, Q^*; \sigma)$ . Fifth,  $\text{URCT}_2(h; \sigma)$  is sharper than  $\text{URCT}_1(h; \sigma)$  for  $\text{RCT}(h, Q^*; \sigma)$ , but requires a mild additional assumption.

As the likelihood is intractable, it is necessary to make a judgement on how to choose  $\sigma$ , because  $\text{IF}(h, Q_{\text{EX}})$  and  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$  are unknown and cannot be easily estimated. Consider two extreme scenarios. The first is the perfect proposal  $q(\theta, \vartheta) = \pi(\vartheta)$ ; in this case, by Corollary 3 and Remark 2,  $\text{RCT}(h, Q; \sigma) = \{2\pi_Z^\sigma(1/\varrho_Z^\sigma) - 1\}/\sigma^2$ , which we denote by  $\text{RCT}(h, Q_\pi; \sigma)$ , is minimized at  $\sigma_{\text{opt}} = 0.92$ . The second scenario involves a very inefficient proposal corresponding to Corollary 2(iv), so that  $\text{RCT}(h, Q^*; \sigma) = \text{LRCT}_2(\sigma)$ , which is minimized at  $\sigma_{\text{opt}} = 1.68$ . If we choose  $\sigma_{\text{opt}} = 1.68$  over  $\sigma_{\text{opt}} = 0.92$  in the first scenario, then  $\text{RCT}(h, Q_\pi; \sigma)$

risks from 5.36 to 12.73. Conversely, if we choose  $\sigma_{\text{opt}} = 0.92$  over  $\sigma_{\text{opt}} = 1.68$  in the second scenario, the relative computing time  $\text{RCT}(h, Q^*; \sigma)$  rises from 1.51 to 2.29. This suggests that the penalty in choosing the wrong value is much more severe if we incorrectly assume we are in the second scenario than if we incorrectly assume we are in the first scenario. This is because as  $\text{IF}(h/Q_{\text{EX}}, \tilde{Q}_{\text{EX}})$  increases,  $\text{LRCT}_2(\sigma)$  is very flat relative to  $\text{RCT}(h, Q_\pi; \sigma)$  as a function of  $\sigma$ . In practice, choosing  $\sigma_{\text{opt}}$  slightly greater than 1.0 seems sensible. For example, a value of  $\sigma = 1.2$  leads to an increase in  $\text{RCT}(h, Q_\pi; \sigma)$  from the minimum value of 5.36 to 6.10 and an increase in  $\text{LRCT}_2(\sigma)$  from the minimum value of 1.51 to 1.75. In Appendix 2, we compute lower and upper bounds for the minimizing argument of  $\text{RCT}(h, Q^*; \sigma)$  for various values of  $\text{IF}(h/Q_{\text{EX}}, \tilde{Q}_{\text{EX}})$ .

Some caution should be exercised in interpreting these results, as the lower bounds apply to  $\text{RCT}(h, Q^*; \sigma)$  but not, in general, to  $\text{RCT}(h, Q; \sigma)$ . Similarly, while  $\text{URCT}_4(h; \sigma)$  and the lower bounds become exact for  $\text{RCT}(h, Q^*; \sigma)$  as  $\text{IF}(h/Q_{\text{EX}}, \tilde{Q}_{\text{EX}}) \rightarrow \infty$ , they only provide upper bounds for  $\text{RCT}(h, Q; \sigma)$ . However, there are scenarios in which  $\text{RCT}(h, Q^*; \sigma) \approx \text{RCT}(h, Q; \sigma)$ . This is the case when the acceptance probability  $\alpha_{\text{EX}}$  of  $Q_{\text{EX}}$  is close to unity so that  $\alpha_{Q^*} \approx \alpha_Q$ , which occurs, for instance, when the proposal is reversible with respect to a distribution approximating  $\pi$  or when the proposal is a random walk with a very small step size.

The numerical results in this section are based on Assumption 2. However, the bounds on the relative inefficiencies of  $Q$  and  $Q^*$  presented in Corollaries 1 and 2 can be calculated for any other noise distribution  $g(z)$ , subject to  $\int \exp(z)g(z) dz = 1$ . For example, one could consider a Laplace distribution with mean  $\log(1 - \sigma^2)$  and scale parameter  $\sigma/\sqrt{2}$  as in [Sherlock et al. \(2015\)](#). These bounds can in turn be used to construct corresponding bounds on the relative computing times of  $Q$  and  $Q^*$ , provided that an appropriate penalization term can be found to relate the parameterization of the noise distribution to the computational effort of obtaining the likelihood estimator.

### 3.6. Discussion

We now compare informally the bound  $\text{LRIF}_2(\sigma) = 1/\{2\Phi(-\sigma/\sqrt{2})\}$  in Corollary 2(iv) with the results in [Sherlock et al. \(2015\)](#). These authors make Assumption 1 and also assume that the target factorizes into  $d$  independent and identically distributed components and that the proposal is an isotropic Gaussian random walk of jump size  $d^{-1/2}l$ . In the Gaussian noise case, for  $h(\theta) = \theta_1$  where  $\theta = (\theta_1, \dots, \theta_d)$ , their results and a standard calculation with their diffusion limit suggest that as  $d \rightarrow \infty$ ,

$$\frac{\text{IF}(h, Q; \sigma, l)}{\text{IF}(h, Q_{\text{EX}}; l)} = \text{RIF}(h, Q; \sigma, l) \rightarrow \text{ARIF}(\sigma, l) = \frac{J_{\sigma^2=0}(l)}{J_{\sigma^2}(l)} = \frac{\Phi(-l/2)}{\Phi\{-(2\sigma^2 + l^2)^{1/2}/2\}},$$

where the expression for  $J_{\sigma^2}(l)$  is given by equations (3.2) and (3.3) in [Sherlock et al. \(2015\)](#). We observe that  $\text{ARIF}(\sigma, l)$  converges to  $\text{LRIF}_2(\sigma)$  as  $l \rightarrow 0$ . This is unsurprising. As  $d \rightarrow \infty$ , we conjecture that in this scenario the conditions of Corollary 2(iv) hold; in particular,  $\text{IF}(h/Q_{\text{EX}}, \tilde{Q}_{\text{EX}}) \rightarrow \infty$  for any  $l > 0$ . Therefore, in this case,  $\text{RIF}(h, Q^*; \sigma, l) \rightarrow \text{LRIF}_2(\sigma)$ . As  $l \rightarrow 0$ , we have informally that  $Q_{\text{EX}}(\theta) \rightarrow 1$ , so it is reasonable to conjecture that  $\text{RIF}(h, Q; \sigma, l)/\text{RIF}(h, Q^*; \sigma, l) \rightarrow 1$ . If one of these limits holds uniformly, then  $\text{ARIF}(\sigma, l) \rightarrow \text{LRIF}_2(\sigma)$ .

## 4. APPLICATION

### 4.1. Stochastic volatility model

In this section we describe a multivariate partially observed diffusion model, which was introduced by [Chernov et al. \(2003\)](#) and discussed in [Huang & Tauchen \(2005\)](#). The logarithm of

regularly observed price  $P(t)$  evolves according to

$$\begin{aligned} d \log P(t) &= \mu_y dt + \sigma(t) dB(t), \\ dv_1(t) &= -k_1\{v_1(t) - \mu_1\} dt + \sigma_1 dW_1(t), \\ dv_2(t) &= -k_2 v_2(t) dt + \{1 + \beta_{12} v_2(t)\} dW_2(t), \end{aligned}$$

with  $\sigma(t) = \text{s-exp}[\{v_1(t) + \beta_2 v_2(t)\}/2]$ , where the function  $\text{s-exp}(\cdot)$  is a spliced exponential function to ensure nonexplosive growth. The leverage parameters corresponding to the correlations between the driving Brownian motions are  $\phi_1 = \text{corr}\{B(t), W_1(t)\}$  and  $\phi_2 = \text{corr}\{B(t), W_2(t)\}$ . The two components for volatility allow for quite sudden changes in the log price while retaining long memory in volatility. We note that the Brownian motion of the price process may be expressed as  $dB(t) = a_1 dW_1(t) + a_2 dW_2(t) + \sqrt{b} d\bar{B}(t)$ , where  $a_1 = \phi_1(1 - \phi_2^2)/(1 - \phi_1^2\phi_2^2)$ ,  $a_2 = \phi_2(1 - \phi_1^2)/(1 - \phi_1^2\phi_2^2)$  and  $b = (1 - \phi_1^2)(1 - \phi_2^2)/(1 - \phi_1^2\phi_2^2)$ . Here  $\bar{B}(t)$  is an independent Brownian motion. Suppose that the log prices are observed at equally spaced times  $\tau_1 < \dots < \tau_T < \tau_{T+1}$ , and let  $\Delta = \tau_{s+1} - \tau_s$  for any  $s$ , which gives returns  $Y_s = \log P(\tau_{s+1}) - \log P(\tau_s)$  for  $s = 1, \dots, T$ . The distribution of these returns conditional on the volatility paths and the driving processes  $W_1(t)$  and  $W_2(t)$  can be expressed in closed form as  $Y_s \sim N(\mu_y \Delta + a_1 Z_{1,s} + a_2 Z_{2,s}, b\sigma_s^{2*})$  where

$$Z_{1,s} = \int_{\tau_s}^{\tau_{s+1}} \sigma(u) dW_1(u), \quad Z_{2,s} = \int_{\tau_s}^{\tau_{s+1}} \sigma(u) dW_2(u), \quad \sigma_s^{2*} = \int_{\tau_s}^{\tau_{s+1}} \sigma^2(u) du. \quad (13)$$

An Euler scheme is used to approximate the evolution of the volatilities  $v_1(t)$  and  $v_2(t)$  by placing  $M - 1$  latent points between  $\tau_s$  and  $\tau_{s+1}$ . The volatility components are denoted by  $v_{1,1}^s, \dots, v_{1,M-1}^s$  and  $v_{2,1}^s, \dots, v_{2,M-1}^s$ . For notational convenience, the start and end points are set to  $v_{1,0}^s = v_1(\tau_s)$  and  $v_{1,M}^s = v_1(\tau_{s+1})$ , and similarly for  $v_2(t)$ . These latent points are evenly spaced in time by  $\delta = \Delta/M$ . The equation for the Euler evolution, starting at  $v_{1,0}^s = v_{1,M}^{s-1}$  and  $v_{2,0}^s = v_{2,M}^{s-1}$ , is

$$\begin{aligned} v_{1,m+1}^s &= v_{1,m}^s - k_1(v_{1,m}^s - \mu_1)\delta + \sigma_1 \sqrt{\delta} u_{1,m}, \\ v_{2,m+1}^s &= v_{2,m}^s - k_2 v_{2,m}^s \delta + (1 + \beta_{12} v_{2,m}^s) \sqrt{\delta} u_{2,m} \quad (m = 0, \dots, M-1), \end{aligned}$$

where  $u_{1,m} \sim N(0, 1)$  and  $u_{2,m} \sim N(0, 1)$ . Conditional on these trajectories and the innovations, the distribution of the returns has a closed form, so that  $Y_s \sim N(\mu_y \Delta + a_1 \hat{Z}_{1,s} + a_2 \hat{Z}_{2,s}, b\hat{\sigma}_s^{2*})$ , where  $\hat{Z}_{1,s}$ ,  $\hat{Z}_{2,s}$  and  $\hat{\sigma}_s^{2*}$  are the Euler approximations to the corresponding expressions in (13).

#### 4.2. Empirical results for the error of the loglikelihood estimator

We consider  $T$  daily returns,  $y = (y_1, \dots, y_T)$ , for the S&P 500 index from three different datasets:  $T = 40$ , 29 April 1983 to 27 June 1983;  $T = 300$ , 27 April 1982 to 27 June 1983 or  $T = 2700$ , 7 March 1975 to 7 November 1985. In this subsection we investigate Assumptions 1 and 2 by examining the behaviour of the error of the loglikelihood estimator  $Z = \log \hat{p}_N(y | \theta) - \log p(y | \theta)$  for the nine-dimensional parameter vector  $\theta = (k_1, \mu_1, \sigma_1, k_2, \beta_{12}, \beta_2, \mu_y, \phi_1, \phi_2)$  of the stochastic volatility model introduced in the previous subsection. The likelihood is estimated using the bootstrap particle filter with  $N$  particles. We select values of  $N$  to ensure that, for each dataset, the variance of  $Z$  evaluated at  $\bar{\theta}$  is approximately unity, where  $\bar{\theta}$  is the posterior mean corresponding to a vague prior. We take  $\delta = 0.5$  for the Euler scheme.

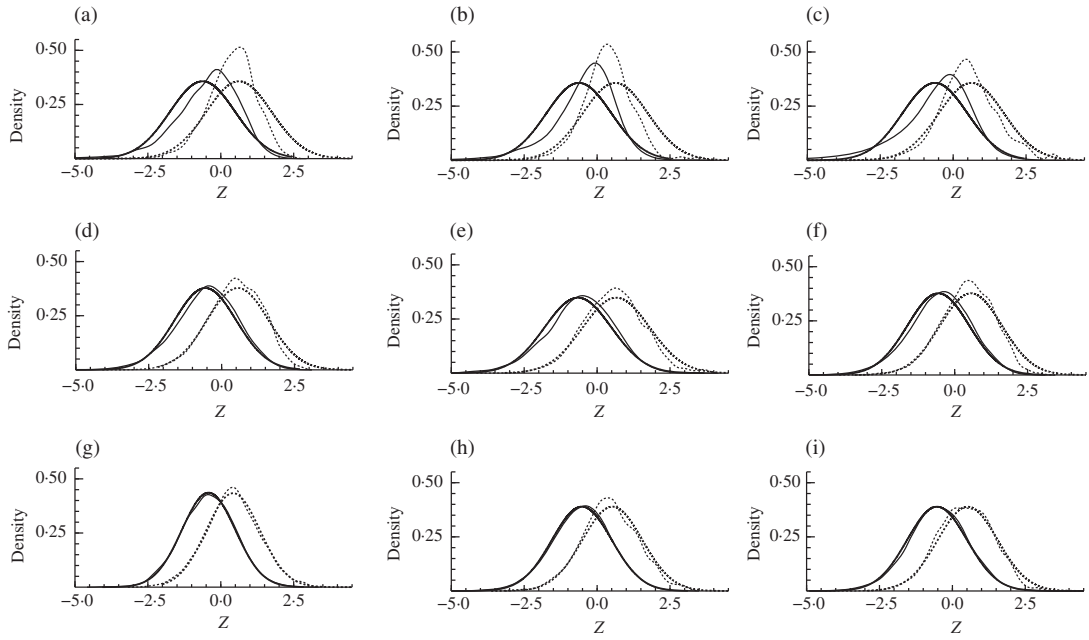


Fig. 2. Estimated and theoretical densities for  $Z$ , in the three cases where  $T = 40$  and  $N = 4$  (panels (a)–(c)),  $T = 300$  and  $N = 80$  (panels (d)–(f)), and  $T = 2700$  and  $N = 700$  (panels (g)–(i)); the kernel density estimates associated with  $g_N(z | \bar{\theta})$  (thin solid) and  $\pi_N(z | \bar{\theta})$  (thin dotted) evaluated at the posterior mean  $\bar{\theta}$  are plotted in panels (a), (d) and (g), the estimates evaluated over values from the posterior  $\pi(\theta)$  are plotted in panels (b), (e) and (h), while those evaluated over values from  $\int \pi(d\theta)q(\vartheta, \theta)$  are plotted in panels (c), (f) and (i); the theoretical densities  $g_Z^\sigma(z)$  (thick solid) and  $\pi_Z^\sigma(z)$  (thick dotted) are overlaid.

Panels (a), (d) and (g) of Fig. 2 display the kernel density estimates corresponding to the density of  $Z$  for  $\theta = \bar{\theta}$ , denoted by  $g_N(z | \bar{\theta})$ , which is obtained by running  $S = 6000$  particle filters at this value. As  $p(y | \bar{\theta})$  is unknown, it is estimated by averaging these estimates. The Metropolis–Hastings algorithm is then used to obtain samples from  $\pi_N(z | \bar{\theta}) = \exp(z)g_N(z | \bar{\theta})$ . On each kernel density estimate we overlay the corresponding assumed density,  $g_Z^\sigma(z)$  or  $\pi_Z^\sigma(z)$ , where  $\sigma^2$  is the sample variance of  $Z$  over the  $S$  particle filters. For  $T = 40$ , there is a discrepancy between the assumed Gaussian densities and the kernel density estimates representing  $g_N(z | \bar{\theta})$  and  $\pi_N(z | \bar{\theta})$ . In particular, although  $g_N(z | \bar{\theta})$  is well approximated over most of its support, it is slightly lighter-tailed than the assumed Gaussian density in the right tail and much heavier-tailed in the left tail. This translates into a smaller discrepancy between  $g_N(z | \bar{\theta})$  and  $\pi_N(z | \bar{\theta})$  and a higher acceptance rate for the pseudo-marginal algorithm than the Gaussian assumption suggests. For  $T = 300$  and  $T = 2700$ , the assumed Gaussian densities are very accurate.

We also examine  $Z$  when  $\theta$  is distributed according to  $\pi(\theta)$ . We record 200 samples from  $\pi(\theta)$ , for  $T = 40, 300$  and  $2700$ . These samples are obtained using the pseudo-marginal algorithm. We use a multivariate Student- $t$  random walk proposal  $q(\vartheta, \theta)$  with step size proportional to  $T^{-1/2}$  on the parameter components transformed to the real line. For each of these samples, we run the particle filter 300 times in order to estimate the true likelihood at these values. The resulting kernel density estimates, corresponding to the densities  $\int \pi(d\theta)g_N(z | \theta)$  and  $\int \pi(d\theta)\pi_N(z | \theta)$ , are displayed in panels (b), (e) and (h) of Fig. 2. We similarly examine the density of  $Z$  when  $\theta$  is distributed according to the marginal proposal density in the stationary regime  $\int \pi(d\theta)q(\vartheta, \theta)$ . Panels (c), (f) and (i) of Fig. 2 show the resulting density estimates. In both scenarios, Assumptions 1 and 2 are problematic for  $T = 40$ , as  $g_N(z | \bar{\theta})$  is not close to Gaussian since



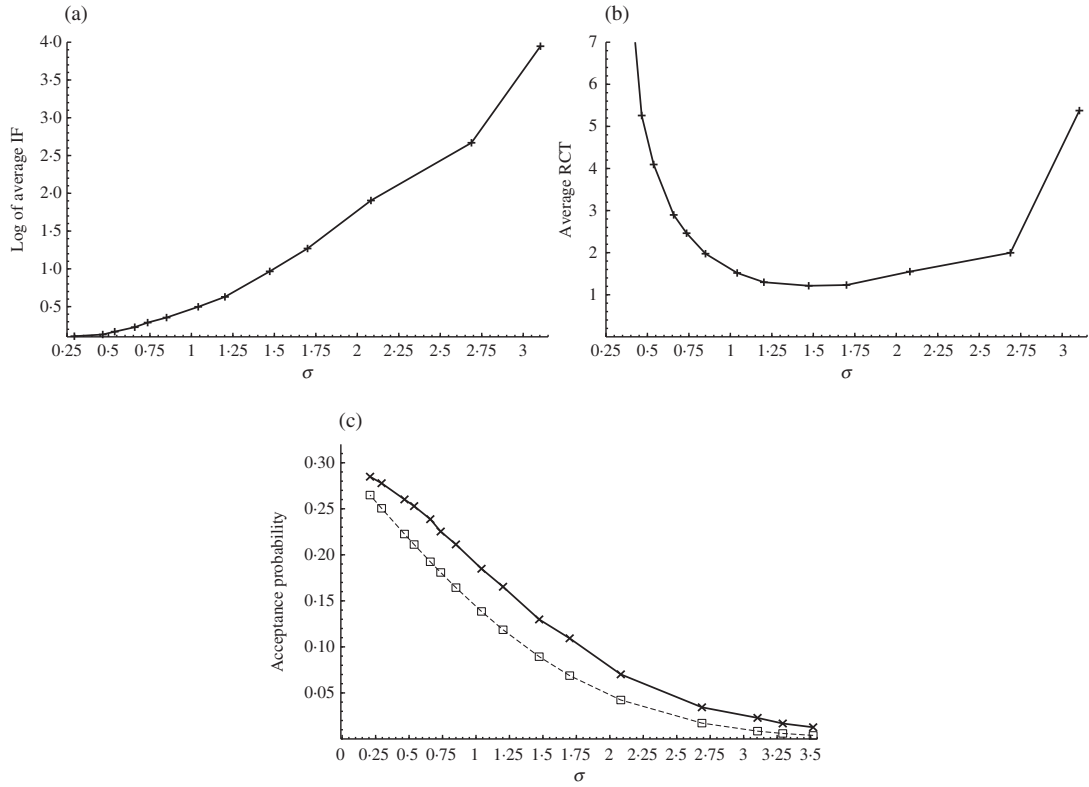


Fig. 3. Performance summary for the pseudo-marginal algorithm applied to the S&P 500 daily returns data for the Huang & Tauchen (2005) two-factor model, with  $T = 300$ : (a) inefficiencies  $\text{IF}$ , averaged over the nine parameter components, plotted against  $\sigma$ ; (b) relative computing times  $\text{RCT}$ , averaged over the nine parameter components, plotted against  $\sigma$ ; (c) marginal acceptance probability  $\pi(\varrho_0)$  (crosses) together with the lower bound  $2\Phi(-\sigma/\sqrt{2})\pi(\varrho_{\text{EX}})$  (squares) plotted against  $\sigma$ .

the central limit theorem provides a poor approximation. Moreover, since  $T$  is small,  $\pi(\theta)$  and  $\int \pi(d\vartheta)q(\vartheta, \theta)$  are relatively diffuse. Consequently,  $g_N(z | \bar{\theta})$  is not close to  $g_N(z | \theta)$  marginalized over  $\pi(\theta)$  or  $\int \pi(d\vartheta)q(\vartheta, \theta)$ . For  $T = 300$  and  $T = 2700$ , the assumed densities  $g_z^\sigma(z)$  and  $\pi_z^\sigma(z)$  are close to the corresponding kernel estimates and Assumptions 1 and 2 appear to capture reasonably well the salient features of the densities associated with  $Z$ . In particular, the approximation suggested by the central limit theorem becomes very good. Additionally,  $\pi(\theta)$  and  $\int \pi(d\vartheta)q(\vartheta, \theta)$  are sufficiently concentrated to ensure that the variance of  $Z$  as a function of  $\theta$  exhibits little variability.

#### 4.3. Empirical results for the pseudo-marginal algorithm

We examine the behaviour of the pseudo-marginal algorithm when sampling from the posterior density  $\pi(\theta)$  with  $\delta = 0.05$  and  $T = 300$  for various values of  $N$ . We use the same pseudo-marginal algorithm as described in the previous subsection.

The standard deviation  $\sigma(\bar{\theta}; N)$  of  $\log \hat{p}_N(y | \bar{\theta})$ , where  $\bar{\theta}$  is the posterior mean, is evaluated by Monte Carlo simulations. For each value of  $N$ , we compute the inefficiencies, denoted by  $\text{IF}$ , and the corresponding approximate relative computing times, denoted by  $\text{RCT}$ , of all parameter components. The quantity  $\text{RCT}$  is computed as  $\text{IF}/\sigma^2(\bar{\theta}; N)$  divided by the inefficiency of  $Q$  when  $N = 2000$ , the latter being an approximation of the inefficiency of  $Q_{\text{EX}}$ . The results are very similar for all parameter components and so, for clarity, Fig. 3 shows the average quantities



over the nine components. For most parameters, the optimal value for  $\sigma(\bar{\theta}; N)$  is between 1.2 and 1.5, corresponding to  $N = 40$  and 60. The results agree with the bound  $\text{URCT}_4(h; \sigma)$  in § 3.5. This can be partly explained by the inefficiencies associated with  $\tilde{Q}$  for  $N = 2000$  being large, suggesting that the inefficiencies associated with  $\tilde{Q}_{\text{EX}}$  are large.

As all the bounds in the paper are based on  $Q^*$ , it is useful to assess the discrepancy between  $Q$  and  $Q^*$ . One approach is to examine the marginal acceptance probability  $\tilde{\pi}(\varrho_Q)$  under  $Q$  against  $\sigma = \sigma(\bar{\theta}, N)$  as  $N$  varies. Using the acceptance criterion (7) of  $Q^*$ , we obtain under Assumptions 1 and 2 that  $\tilde{\pi}(\varrho_Q) \geq 2\Phi(-\sigma/\sqrt{2})\pi(\varrho_{\text{EX}})$ . If  $Q$  and  $Q^*$  are close in the sense of having similar marginal acceptance probabilities, then we expect  $\tilde{\pi}(\varrho_Q)$  to have a similar shape to its lower bound where  $\pi(\varrho_{\text{EX}})$  is approximated using  $\tilde{\pi}(\varrho_Q)$  with  $N = 2000$ . For this model, the two functions on either side of the inequality, displayed in Fig. 3, are similar.

#### ACKNOWLEDGEMENT

The authors thank the editor, associate editor and reviewers for their comments, which have helped to improve the paper significantly. In particular, we are grateful to one of the reviewers for suggesting the proof of uniform ergodicity in Lemma 3. Arnaud Doucet was partially supported by the U.K. Engineering and Physical Sciences Research Council, and Robert Kohn by the Australian Research Council.

#### SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes the proofs of Propositions 2–4 and Lemmas 1 and 3, as well as an additional example.

#### APPENDIX 1

*Proof of Lemma 2.* It is straightforward to establish that  $Q^*$  is  $\tilde{\pi}$ -reversible. Moreover, for any  $a, b \geq 0$ ,  $\min(1, a) \min(1, b) \leq \min(1, ab)$ , so  $\alpha_{Q^*}\{(\theta, z), (\vartheta, w)\} \leq \alpha_Q\{(\theta, z), (\vartheta, w)\}$  for any  $\theta, z, \vartheta$  and  $w$ . Hence, Theorem 4 in Tierney (1998), which is a general state-space extension of Peskun (1973), applies and yields the result.  $\square$

*Proof of Theorem 1.* Without loss of generality, let  $h \in L_0^2(\Theta, \pi)$ . By Theorem 7 of Andrieu & Vihola (2015),  $\text{IF}(h, Q_{\text{EX}}) \leq \text{IF}(h, Q)$ , and by Lemma 2,  $\text{IF}(h, Q) \leq \text{IF}(h, Q^*)$ , where  $\text{IF}(h, Q^*) < \infty$  by assumption. Hence  $\text{IF}(h, Q_{\text{EX}}) < \infty$ , and Proposition 2 applied to  $Q_{\text{EX}}$  yields that  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}), \tilde{\pi}(h^2/\varrho_{\text{EX}}^2) < \infty$  and

$$\pi(h^2)\{1 + \text{IF}(h, Q_{\text{EX}})\} = \pi(\varrho_{\text{EX}})\tilde{\pi}(h^2/\varrho_{\text{EX}}^2)\{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}. \quad (\text{A1})$$

Since the assumptions of Lemma 3 are satisfied, we can substitute (A1) into (11) to obtain

$$1 + \text{IF}(h, Q^*) = \pi_z(1/\varrho_z) \frac{\{1 + \text{IF}(h, Q_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \left[ 1 + \text{IF}\{h/(\varrho_{\text{EX}}\varrho_z), \tilde{Q}^*\} \right]. \quad (\text{A2})$$

We now provide a spectral representation for  $\text{IF}\{h/(\varrho_{\text{EX}}\varrho_z), \tilde{Q}^*\}$ . With  $\tilde{\pi} \otimes \tilde{\pi}_z(d\theta, dz) = \tilde{\pi}(d\theta)\tilde{\pi}_z(dz)$ ,

$$\begin{aligned} \text{IF}\{h/(\varrho_{\text{EX}}\varrho_z), \tilde{Q}^*\} &= 1 + 2 \sum_{n=1}^{\infty} \frac{\langle \varrho_{\text{EX}}^{-1} \varrho_z^{-1} h, (\tilde{Q}^*)^n \varrho_{\text{EX}}^{-1} \varrho_z^{-1} h \rangle_{\tilde{\pi} \otimes \tilde{\pi}_z}}{\tilde{\pi} \otimes \tilde{\pi}_z(\varrho_z^{-2} \varrho_{\text{EX}}^{-2} h^2)} \\ &= 1 + 2 \sum_{n=1}^{\infty} \frac{\langle \varrho_z^{-1}, (\tilde{Q}_z)^n \varrho_z^{-1} \rangle_{\tilde{\pi}_z} \langle \varrho_{\text{EX}}^{-1} h, (\tilde{Q}_{\text{EX}})^n \varrho_{\text{EX}}^{-1} h \rangle_{\tilde{\pi}}}{\tilde{\pi}_z(\varrho_z^{-2}) \tilde{\pi}(\varrho_{\text{EX}}^{-2} h^2)} \end{aligned} \quad (\text{A3})$$

and, as  $\tilde{Q}_Z$  and  $\tilde{Q}_{EX}$  are reversible, the following spectral representations, as in (8), hold:

$$\begin{aligned}\phi_n(1/\varrho_Z, \tilde{Q}_Z) &= \frac{\langle \varrho_Z^{-1}, (\tilde{Q}_Z)^n \varrho_Z^{-1} \rangle_{\tilde{\pi}_Z} - \{\tilde{\pi}_Z(\varrho_Z^{-1})\}^2}{\mathbb{V}_{\tilde{\pi}_Z}(\varrho_Z^{-1})} = \int_{-1}^1 \lambda^n \tilde{e}_Z(d\lambda), \\ \phi_n(h/\varrho_{EX}, \tilde{Q}_{EX}) &= \frac{\langle \varrho_{EX}^{-1} h, (\tilde{Q}_{EX})^n \varrho_{EX}^{-1} h \rangle_{\tilde{\pi}}}{\tilde{\pi}(\varrho_{EX}^{-2} h^2)} = \int_{-1}^1 \omega^n \tilde{e}_{EX}(d\omega),\end{aligned}$$

where, to simplify notation, we define

$$\mathbb{V}_{\tilde{\pi}_Z}(\varrho_Z^{-1}) = \tilde{\pi}_Z[\{\varrho_Z^{-1} - \tilde{\pi}_Z(\varrho_Z^{-1})\}^2], \quad \tilde{e}_Z(d\lambda) = e(\varrho_Z^{-1}, \tilde{Q}_Z)(d\lambda), \quad \tilde{e}_{EX}(d\omega) = e(\varrho_{EX}^{-1} h, \tilde{Q}_{EX})(d\omega).$$

Using  $\tilde{\pi}_Z(\varrho_Z^{-1}) = 1/\pi_Z(\varrho_Z)$ , we can rewrite  $\text{IF}\{h/(\varrho_{EX}\varrho_Z), \tilde{Q}^*\}$  in (A3) as

$$\begin{aligned}1 + 2 \sum_{n=1}^{\infty} \frac{1}{\tilde{\pi}_Z(\varrho_Z^{-2})} \left\{ \mathbb{V}_{\tilde{\pi}_Z}(\varrho_Z^{-1}) \int \lambda^n \tilde{e}_Z(d\lambda) + \frac{1}{\pi_Z(\varrho_Z)^2} \right\} \int \omega^n \tilde{e}_{EX}(d\omega) \\ = 1 + 2(1 - \gamma) \int \frac{\omega}{1 - \omega} \tilde{e}_{EX}(d\omega) + 2\gamma \iint \frac{\lambda\omega}{1 - \lambda\omega} \tilde{e}_Z(d\lambda) \tilde{e}_{EX}(d\omega) \\ = -1 + 2(1 - \gamma) \int \left(1 + \frac{\omega}{1 - \omega}\right) \tilde{e}_{EX}(d\omega) + 2\gamma \iint \left(1 + \frac{\omega\lambda}{1 - \lambda\omega}\right) \tilde{e}_Z(d\lambda) \tilde{e}_{EX}(d\omega),\end{aligned}\quad (\text{A4})$$

where the second expression is finite since  $\int (1 + \omega)(1 - \omega)^{-1} \tilde{e}_{EX}(d\omega) = \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX}) < \infty$  and

$$\gamma = \mathbb{V}_{\tilde{\pi}_Z}(\varrho_Z^{-1})/\tilde{\pi}_Z(\varrho_Z^{-2}) = \{\pi_Z(\varrho_Z^{-1}) - 1/\pi_Z(\varrho_Z)\}/\pi_Z(\varrho_Z^{-1}). \quad (\text{A5})$$

Upon rearranging (A4), we have

$$\begin{aligned}1 + \text{IF}\{h/(\varrho_{EX}\varrho_Z), \tilde{Q}^*\} &= \{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})\}(1 - \gamma) + \gamma\beta \\ &= \frac{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})}{\pi_Z(\varrho_Z)\pi_Z(\varrho_Z^{-1})} + \left\{ \frac{\pi_Z(\varrho_Z^{-1}) - 1/\pi_Z(\varrho_Z)}{\pi_Z(\varrho_Z^{-1})} \right\} \beta,\end{aligned}\quad (\text{A6})$$

with

$$\frac{\beta}{2} = \iint \frac{\tilde{e}_Z(d\lambda) \tilde{e}_{EX}(d\omega)}{1 - \omega\lambda} = \sum_{n=0}^{\infty} \phi_n(h/\varrho_{EX}, \tilde{Q}_{EX}) \phi_n(1/\varrho_Z, \tilde{Q}_Z). \quad (\text{A7})$$

By substituting (A6) into (A2) and using that

$$\begin{aligned}\text{IF}(h, Q^*) &= \pi_Z(1/\varrho_Z) \frac{1 + \text{IF}(h, Q_{EX})}{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})} \left[ 1 + \text{IF}\{h/(\varrho_{EX}\varrho_Z), \tilde{Q}^*\} \right] - 1 \\ &= \frac{1 + \text{IF}(h, Q_{EX})}{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})} \left\{ \pi_Z(\varrho_Z^{-1}) - \frac{1}{\pi_Z(\varrho_Z)} \right\} \beta + \frac{1 + \text{IF}(h, Q_{EX})}{\pi_Z(\varrho_Z)} - 1,\end{aligned}\quad (\text{A8})$$

we obtain the result.  $\square$

*Proof of Corollary 1.* Upon dividing (A8) by  $\text{IF}(h, Q_{EX})$ , we obtain

$$\text{RIF}(h, Q^*) = \frac{\pi_Z(\varrho_Z^{-1})\{1 + \text{IF}(h, Q_{EX})\}}{\text{IF}(h, Q_{EX})\{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})\}} A - \frac{1}{\text{IF}(h, Q_{EX})}, \quad (\text{A9})$$

where  $A$  is the quantity in (A6) and can be expressed in terms of  $\gamma$ , defined in (A5), as

$$\begin{aligned} A &= 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 2\gamma \iint \left\{ \frac{1}{(1-\omega)} - \frac{1}{(1-\lambda\omega)} \right\} \tilde{e}_z(d\lambda) \tilde{e}_{\text{EX}}(d\omega) \\ &= 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 2\gamma \iint \frac{\omega(1-\lambda)}{(1-\omega)(1-\omega\lambda)} \tilde{e}_z(d\lambda) \tilde{e}_{\text{EX}}(d\omega). \end{aligned}$$

Lemma 3 ensures that the kernel  $\tilde{Q}_z$  is positive, implying that  $\tilde{e}_z\{[0, 1]\} = 1$ . Hence,

$$\iint \left\{ \frac{\omega(1-\lambda)}{(1-\omega)(1-\omega\lambda)} - \frac{\omega(1-\lambda)}{(1-\omega)} \right\} \tilde{e}_z(d\lambda) \tilde{e}_{\text{EX}}(d\omega) = \iint \frac{\omega^2(1-\lambda)\lambda}{(1-\omega)(1-\omega\lambda)} \tilde{e}_z(d\lambda) \tilde{e}_{\text{EX}}(d\omega) \geq 0.$$

We can now bound  $A$  from above as follows:

$$\begin{aligned} A &\leq 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 2\gamma \iint \frac{\omega(1-\lambda)}{(1-\omega)} \tilde{e}_z(d\lambda) \tilde{e}_{\text{EX}}(d\omega) \\ &= 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - \gamma \left\{ 1 - \int \lambda \tilde{e}_z(d\lambda) \right\} \int \frac{2\omega}{(1-\omega)} \tilde{e}_{\text{EX}}(d\omega) \\ &= 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - \gamma(1 - \phi_z) \{ \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 1 \} \\ &= \{ 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \} \{ \phi_z + (1 - \phi_z)(1 - \gamma) \} + 2(1 - \phi_z)\gamma \\ &\leq \{ 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \} \{ \phi_z + (1 - \phi_z)(1 - \gamma) + 2(1 - \phi_z)\gamma \} \\ &= \{ 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \} \left\{ 2(1 - \phi_z/2) - \frac{(1 - \phi_z)}{\pi_z(1/\varrho_z)\pi_z(\varrho_z)} \right\}, \end{aligned} \quad (\text{A10})$$

where we have used the identity  $\phi_z = \int \lambda \tilde{e}_z(d\lambda)$ . The last inequality is established by noting that  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$  and  $\gamma$  are nonnegative. Substituting the expression into (A9) establishes statement (i) of the corollary. By observing that if  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \geq 1$ , then (A10) is bounded above by

$$\begin{aligned} &\{ 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \} \left[ \phi_z + \frac{(1 - \phi_z)}{\pi_z(1/\varrho_z)\pi_z(\varrho_z)} + (1 - \phi_z) \left\{ 1 - \frac{1}{\pi_z(1/\varrho_z)\pi_z(\varrho_z)} \right\} \right] \\ &= 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}), \end{aligned}$$

we prove the inequality in (ii). □

*Proof of Corollary 2.* We establish the upper bound  $\text{URIF}_3(h)$  in (i) by noting that (A10) implies

$$A \leq \{ 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \} \{ \phi_z + (1 - \phi_z)(1 - \gamma) \} + 2(1 - \phi_z)\gamma,$$

where  $A$  is the quantity in (A6) and  $\gamma$  is given by (A5). Upon substituting into (A9), we obtain

$$\begin{aligned} \text{RIF}(h, Q^*) + \frac{1}{\text{IF}(h, Q_{\text{EX}})} &\leq \frac{\{ 1 + \text{IF}(h, Q_{\text{EX}}) \}}{\text{IF}(h, Q_{\text{EX}})} \left\{ \phi_z \pi_z(\varrho_z^{-1}) + \frac{(1 - \phi_z)}{\pi_z(\varrho_z)} \right\} \\ &\quad + \frac{2(1 - \phi_z) \{ 1 + \text{IF}(h, Q_{\text{EX}}) \}}{\text{IF}(h, Q_{\text{EX}}) \{ 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \}} \left\{ \pi_z(\varrho_z^{-1}) - \frac{1}{\pi_z(\varrho_z)} \right\}, \end{aligned}$$

and, after further manipulations, we have

$$\begin{aligned} \text{RIF}(h, Q^*) &\leq \phi_z \{ \pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z) \} + 1/\pi_z(\varrho_z) \\ &\quad + \frac{1}{\text{IF}(h, Q_{\text{EX}})} \left[ \phi_z \{ \pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z) \} + \frac{1}{\pi_z(\varrho_z)} - 1 \right] \\ &\quad + 2 \frac{\{ 1 + \text{IF}(h, Q_{\text{EX}}) \}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \{ \pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z) \} (1 - \phi_z) \end{aligned}$$

$$\begin{aligned}
&\leq \phi_z \{ \pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z) \} + 1/\pi_z(\varrho_z) \\
&\quad + \frac{1}{\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \left[ \phi_z \{ \pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z) \} + \frac{1}{\pi_z(\varrho_z)} - 1 \right] \\
&\quad + \frac{2}{\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \{ \pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z) \} (1 - \phi_z),
\end{aligned}$$

as  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \leq \text{IF}(h, Q_{\text{EX}})$  from Proposition 2.

To establish the upper bound  $\text{URIF}_4(h)$  in (ii), we use that, in the right-hand side of the equality (A9), the term  $\beta$  defined in (A7) and appearing in  $\mathcal{A}$  satisfies the inequality

$$\beta = \iint \frac{2}{(1-\lambda\omega)} \tilde{e}_z(d\lambda) \tilde{e}_{\text{EX}}(d\omega) \leq \int \frac{2}{(1-\lambda)} \tilde{e}_z(d\lambda) = 1 + \text{IF}(1/\varrho_z, \tilde{Q}_z), \quad (\text{A11})$$

where  $\text{IF}(1/\varrho_z, \tilde{Q}_z) = \int (1+\lambda)/(1-\lambda) \tilde{e}_z(d\lambda) < \infty$  by assumption. Therefore, upon substituting into (A9), we obtain

$$\begin{aligned}
\text{RIF}(h, Q^*) &\leq \frac{\pi_z(\varrho_z^{-1}) \{1 + \text{IF}(h, Q_{\text{EX}})\}}{\text{IF}(h, Q_{\text{EX}}) \{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}} \left\{ 1 - \frac{1}{\pi_z(\varrho_z^{-1}) \pi_z(\varrho_z)} \right\} \{1 + \text{IF}(1/\varrho_z, \tilde{Q}_z)\} \\
&\quad + \frac{\{1 + \text{IF}(h, Q_{\text{EX}})\}}{\text{IF}(h, Q_{\text{EX}})} \frac{1}{\pi_z(\varrho_z)} - \frac{1}{\text{IF}(h, Q_{\text{EX}})} \\
&= \frac{1}{\pi_z(\varrho_z)} + \frac{\{1 + 1/\text{IF}(h, Q_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \{ \pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z) \} \{1 + \text{IF}(1/\varrho_z, \tilde{Q}_z)\} \\
&\quad + \frac{1}{\text{IF}(h, Q_{\text{EX}})} \left\{ \frac{1}{\pi_z(\varrho_z)} - 1 \right\} \\
&\leq \frac{1}{\pi_z(\varrho_z)} + \frac{\{1 + 1/\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \{ \pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z) \} \{1 + \text{IF}(1/\varrho_z, \tilde{Q}_z)\} \\
&\quad + \frac{1}{\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \left\{ \frac{1}{\pi_z(\varrho_z)} - 1 \right\},
\end{aligned}$$

since  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \leq \text{IF}(h, Q_{\text{EX}})$ .

To establish the inequality in (iii), we combine (A6) and (A9) to obtain

$$\begin{aligned}
\text{RIF}(h, Q^*) &= \frac{\pi_z(\varrho_z^{-1}) \{1 + 1/\text{IF}(h, Q_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \gamma \beta + (1 - \gamma) \pi_z(\varrho_z^{-1}) \{1 + 1/\text{IF}(h, Q_{\text{EX}})\} - \frac{1}{\text{IF}(h, Q_{\text{EX}})} \\
&= \frac{1}{\pi_z(\varrho_z)} + \frac{\{1 + 1/\text{IF}(h, Q_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \{ \pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z) \} \beta + \frac{\{1/\pi_z(\varrho_z) - 1\}}{\text{IF}(h, Q_{\text{EX}})} \\
&\geq \frac{1}{\pi_z(\varrho_z)} + \frac{2\{1 + 1/\text{IF}(h, Q_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \{ \pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z) \} + \frac{\{1/\pi_z(\varrho_z) - 1\}}{\text{IF}(h, Q_{\text{EX}})} \quad (\text{A12}) \\
&\geq \frac{1}{\pi_z(\varrho_z)} + \frac{2}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})} \{ \pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z) \}.
\end{aligned}$$

The first inequality follows because the identity for  $\beta$  given in (A11) shows that  $\beta \geq 2$  when  $\tilde{Q}_{\text{EX}}$  is positive. The second inequality follows from  $\text{IF}(h, Q_{\text{EX}}) \geq 0$ .

From (A9), we have that  $\text{RIF}(h, Q^*) \geq 1/\pi_z(\varrho_z)$ , as the second and third terms on the left-hand side of the inequality (A12) are both positive. This establishes the inequality of (iv). We examine the limit of  $\text{RIF}(h, Q^*)$  as  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \rightarrow \infty$ , again noting that  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \leq \text{IF}(h, Q_{\text{EX}})$ . Using the inequality for  $\beta$  in (A11) and the fact that  $\text{IF}(1/\varrho_z, \tilde{Q}_z) < \infty$  by Lemma 3, we obtain the limiting form, as  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \rightarrow \infty$ , given by (12) for  $\text{RIF}(h, Q^*)$ .  $\square$

Table A1. Sandwiching results for different values of  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$ , based on the upper bounds for  $\text{RCT}(h, Q^*; \sigma)$  given by  $\text{URCT}_3(h; \sigma)$  and  $\text{URCT}_4(h; \sigma)$  and on the lower bound  $\text{LRCT}_1(h; \sigma)$

$\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$	1	10	25	100	1000
$\text{RCT}(h, Q^*; \sigma_{\text{opt}})$	(3.201, 5.327)	(2.020, 2.256)	(1.773, 1.876)	(1.595, 1.625)	(1.518, 1.522)
$\sigma_{\text{opt}}$	(0.548, 1.572)	(1.018, 1.598)	(1.205, 1.658)	(1.421, 1.730)	(1.607, 1.730)

## APPENDIX 2

We exploit the two upper bounds  $\text{URCT}_3(h; \sigma)$  and  $\text{URCT}_4(h; \sigma)$ , together with the lower bound  $\text{LRCT}_1(h; \sigma)$ , in order to find an interval where the optimal value  $\sigma_{\text{opt}}$  for  $\text{RCT}(h, Q^*; \sigma)$  lies. We consider how this interval varies as  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$  increases. To do this, we compute the interval where  $\text{LRCT}_1(h; \sigma)$  lies below the minimum of  $\inf_{\sigma} \text{URCT}_3(h; \sigma)$  and  $\inf_{\sigma} \text{URCT}_4(h; \sigma)$ . Table A1 displays this interval together with the minimum of the two upper bounds and the minimum of the lower bound. It is straightforward to see that  $\sigma_{\text{opt}}$  is contained in this interval and  $\text{RCT}(h, Q^*; \sigma_{\text{opt}})$  is contained in the corresponding interval in Table A1. It is apparent that the intervals tighten as  $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$  increases. Similarly, the endpoints of the interval containing  $\text{RCT}(h, Q^*; \sigma_{\text{opt}})$  both decrease, whereas the lower endpoint of the interval containing  $\sigma_{\text{opt}}$  increases. A degree of caution should be exercised, however, as this argument applies to  $Q^*$  rather than  $Q$ .

## REFERENCES

- ANDRIEU, C., DOUCET, A. & HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods (with Discussion). *J. R. Statist. Soc. B* **72**, 269–342.
- ANDRIEU, C. & ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37**, 697–725.
- ANDRIEU, C. & VIHOLA, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Prob.* to appear.
- BAXENDALE, P. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Prob.* **15**, 700–38.
- BEAUMONT, M. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**, 1139–60.
- BÉRARD, J., DEL MORAL, P. & DOUCET, A. (2014). A lognormal central limit theorem for particle approximations of normalizing constants. *Electron. J. Prob.* **19**, 1–28.
- CHERNOV, M., GALLANT, A. R., GHYSSELS, E. & TAUCHEN, G. (2003). Alternative models of stock price dynamics. *J. Economet.* **116**, 225–57.
- DOUC, R. & ROBERT, C. P. (2011). A vanilla Rao–Blackwellization of Metropolis–Hastings algorithms. *Ann. Statist.* **39**, 261–77.
- GEYER, C. J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* **7**, 473–83.
- HÄGGSTRÖM, O. & ROSENTHAL, J. S. (2007). On variance conditions for Markov chain central limit theorems. *Electron. Commun. Prob.* **12**, 454–64.
- HUANG, X. & TAUCHEN, G. (2005). The relative contribution of jumps to total price variation. *J. Finan. Economet.* **3**, 456–99.
- KIPNIS, C. & VARADHAN, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun. Math. Phys.* **104**, 1–19.
- LIN, L., LIU, K. F. & SLOAN, J. (2000). A noisy Monte Carlo algorithm. *Phys. Rev. D* **61**, article no. 074505.
- PESKUN, P. H. (1973). Optimum Monte–Carlo sampling using Markov chains. *Biometrika* **60**, 607–12.
- PITT, M. K., SILVA, R., GIORDANI, P. & KOHN, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *J. Economet.* **171**, 134–51.
- ROBERTS, G. O. & TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, 95–110.
- RUDOLF, D. & ULLRICH, M. (2013). Positivity of hit-and-run and related algorithms. *Electron. Commun. Prob.* **18**, 1–8.
- SHERLOCK, C., THIERY, A. H., ROBERTS, G. O. & ROSENTHAL, J. S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43**, 238–75.
- TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with Discussion). *Ann. Statist.* **21**, 1701–62.
- TIERNEY, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Prob.* **8**, 1–9.

[Received October 2012. Revised November 2014]